

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

In Re: OpenAI, Inc. Copyright Infringement
Litigation

This Document Relates To:

Case No. 1:23-cv-08292
Case No. 1:23-cv-10211
Case No. 1:23-cv-00084
Case No. 1:25-cv-03291
Case No. 1:25-cv-03297
Case No. 1:25-cv-03482
Case No. 1:25-cv-03843

Case No. 1:25-md-3143 (SHS) (OTW)

**DEFENDANT MICROSOFT CORPORATION'S MEMORANDUM OF LAW IN
SUPPORT OF ITS MOTION TO STRIKE PORTIONS OF THE CONSOLIDATED
CLASS ACTION COMPLAINT**

TABLE OF CONTENTS

INTRODUCTION.....	1
FACTUAL AND PROCEDURAL BACKGROUND	2
A. The Scope Of The Existing NY Consolidated Class Cases Against Microsoft.....	2
B. This Court’s Orders.....	4
C. The Consolidated Class Complaint Expands The Scope Of Class Plaintiffs’ Claims Against Microsoft	5
D. Class Plaintiffs File Two New Complaints Against Microsoft	7
ARGUMENT	9
I. PORTIONS OF THE CCAC SHOULD BE STRICKEN BECAUSE THEY EXCEED THE SCOPE OF THE CASE SET BY THE COURT	11
A. Allegations Implicating Non-OpenAI-Trained Models Should Be Stricken Because Only OpenAI-Trained Models Have Been The Subject Of Discovery	12
B. Allegations Regarding GPT Models Released Before GPT-3 And After GPT-4 Turbo, And The Products Those Models Underlie, Should Be Stricken Because They Are Not Within The Scope Of Existing Discovery	14
C. Allegations Regarding Non-Books Should Be Stricken Because The Litigation Has Focused Solely On The Use Of Books For Training OpenAI’s LLMs	17
CONCLUSION	18

TABLE OF AUTHORITIES**Page(s)****Cases**

<i>In re Agent Orange Prod. Liab. Litig.</i> , 597 F. Supp. 740 (E.D.N.Y. 1984)	10
<i>Cummings v. City of New York</i> , No. 19-CV-7723 (CM)(OTW), 2021 WL 1163654 (S.D.N.Y. Mar. 26, 2021)	9
<i>Huebner v. Midland Credit Mgmt., Inc.</i> , 897 F.3d 42 (2d Cir. 2018).....	10
<i>Mahoney v. Yamaha Motor Corp. U.S.A.</i> , 290 F.R.D. 363 (E.D.N.Y. 2013)	10
<i>Palm Beach Strategic Income, LP v. Salzman</i> , 457 F. App'x 40 (2d Cir. 2012)	9, 12
<i>In re Phenylpropanolamine (PPA) Prods. Liab. Litig.</i> , 460 F.3d 1217 (9th Cir. 2006)	10
<i>Starostenko v. UBS AG</i> , No. 19 CIV. 9993 (KPF), 2022 WL 1082533 (S.D.N.Y. Apr. 7, 2022).....	9, 10, 16
<i>Sussman Sales Co., Inc. v. VWR Int'l, LLC</i> , No. 20 CIV. 2869 (JHR), 2025 WL 1004729 (S.D.N.Y. Apr. 1, 2025)	9, 12, 16
<i>In re Worldcom, Inc. Sec. Litig.</i> , No. 03 CIV. 6225, 2004 WL 2889974 (S.D.N.Y. Dec. 15, 2004).....	10, 12, 16

Statutes

17 U.S.C. § 106.....	5
28 U.S.C. § 1407(a)	9, 10

Other Authorities

Fed. R. Civ. P. 12(f).....	9
Fed. R. Civ. P. 16(f).....	9, 10

INTRODUCTION

Plaintiffs' Consolidated Class Action Complaint ("CCAC") expands the scope of the case by adding new allegations (1) that Microsoft infringed their copyrights when training its own models, and (2) that Microsoft and OpenAI infringed their copyrights when training new models developed long after Plaintiffs filed their original lawsuits. Plaintiffs seek to add these allegations after almost two years of litigation and 18 months of discovery. They do not add them at this late date because the allegations have already been the subject of discovery. To the contrary, adding these allegations would balloon the scope of new discovery just months before the Court's deadline to substantially complete document production. The Court should strike these improper new allegations.

If these arguments sound familiar, it is because the Court already ruled on them. The parties previously addressed whether to expand the class claims during the Court's May 22 hearing. Microsoft was then a defendant in the class action cases filed in the Southern District of New York but not in the cases filed in the Northern District of California. At the time of the hearing, Plaintiffs in both sets of cases had pending motions to amend their complaints, seeking to add new allegations about the models and classes at issue. During the hearing, Microsoft argued that the new case consolidating the New York and California cases should proceed with the overlapping theories of liability and models already at issue in the existing New York cases, and should not expand the scope of the case with new allegations that would delay a resolution of the ultimate fair use issue. Ultimately, the Court ordered that the consolidated complaint should include "***no new products or causes of action; only those that have been already asserted in these cases.*** None of the new ones on the proposed amended complaints; just the causes of action that are already existing and the products that are already existing." 5/22/25

Hr'g. Tr. (ECF 80-1) at 20:19-23.¹

The CCAC filed by Plaintiffs disregards the Court's Order. First, it redefines the putative class and broadens Microsoft's alleged infringement to include new claims for any downloading of Plaintiffs' works without apparent limitation, which they seek to use to expand the scope of discovery into Microsoft's search engine business generally and Microsoft's training of LLMs that have nothing to do with OpenAI. Second, the CCAC includes new allegations and new claims about additional OpenAI models and the Microsoft products that use them. These new allegations were the subject of the motions for leave to amend the existing pleading that the Court previously rejected, ordering Plaintiffs to file new cases (which they did). Third, the new putative class definition sweeps in non-book texts even though the litigation has focused solely on the alleged use of books for training OpenAI's LLMs. The CCAC fails to heed the Court's rulings in material respects and the offending portions should be stricken.

FACTUAL AND PROCEDURAL BACKGROUND

A. The Scope Of The Existing NY Consolidated Class Cases Against Microsoft

This MDL consolidated a dozen actions filed in the Southern District of New York and Northern District of California, all of which alleged copyright infringement against Microsoft and/or OpenAI *based solely on the training of specific OpenAI GPT models*. ECF 1 at 1. One set of consolidated cases from New York and one from California were class actions, alleging putative classes comprised of book authors alleging infringement based on the use of their books to train the GPT models.² *Id.* at 2. At the time the MDL was ordered, Microsoft was a defendant

¹ Unless otherwise indicated, ECF numbers referenced herein are to the MDL docket: *In re OpenAI, Inc. Copyright Infringement*, 25-md-03143.

² These cases include *Authors Guild et al. v. OpenAI, Inc., et al.*, No. 1:23-cv-08292-SHS (S.D.N.Y.) ("AG"), *Alter et al. v. OpenAI, Inc., et al.*, No. 1:23-cv-10211-SHS (S.D.N.Y.), and

in only the consolidated NY Class Action.

Consistent with the organizing principle of the MDL, the previous 18 months of discovery in the cases pending before this Court involved interrogatories and document production related to only *OpenAI's GPT models*, including the model training and related products. No discovery was taken with respect to any Microsoft-trained LLM because no such LLM was at issue in either class action. *See, e.g., AG*, ECF 67 (“NY Class Complaint”) ¶¶ 70, 74, 76, 96-101 (allegations solely related to the training of OpenAI’s models); *In re Open AI*, ECF 120 (“CA Class Complaint”) ¶¶ 33-44, (same).

These discovery limitations were consistent with Magistrate Judge Wang’s orders resolving disputes over the scope of discovery. Specifically with respect to the NY Class Action—the only class case to which Microsoft was a party prior to the creation of the MDL—discovery was explicitly limited with respect to the Microsoft products at issue. Only those products powered by GPT-3, GPT-3.5, GPT-4, or GPT-4 Turbo have been the subject of discovery from Microsoft. In fact, Magistrate Judge Wang expressly denied motions seeking discovery from Microsoft into models and products that were later included in the CCAC. In the NY Class Action, Plaintiffs sought to compel a response to a request seeking documents showing the training sets Microsoft used to train its LLMs and unexecuted agreements, arguing that such documents were within the scope of its direct infringement claim against Microsoft. *See AG*, ECF 269. Microsoft argued that the training of Microsoft’s LLMs and the products that use them were irrelevant to Plaintiffs’ claims, which focused on the training of OpenAI’s LLMs. *See id.*, ECF 279. The Court agreed with Microsoft. *See id.*, ECF 293. Similarly, the newspaper

Basbanes, et al. v. Microsoft Corporation, et al., No. 1:24-cv-00084-SHS (S.D.N.Y.), which were consolidated (together, “NY Class Action”); and *In re OpenAI ChatGPT Litigation*, No. 23-cv-3223-AMO (N.D. Cal.) (“*In re OpenAI*” or “CA Class Action”).

Plaintiffs sought to compel discovery into newer-released and future-released Microsoft products. *The New York Times Company v. Microsoft Corp., et al.*, No. 23-cv-11195-SHS (“*NYT*”), ECF 279. In opposition, Microsoft noted that these newer products fell outside the scope of the litigation since they were released well after Plaintiffs filed suit. *Id.*, ECF 294. The Court again sided with Microsoft and denied this new discovery. *Id.*, ECF 365 at 5.

B. This Court’s Orders

To organize the MDL with respect to the multiple pending class actions, the Court ordered Class Plaintiffs to file a single consolidated Class Action Complaint which was “to include *only the same products and causes of action that have already been asserted* in the pending putative class actions.” ECF 60 at 2 (emphasis added). During the initial Case Management Conference, the Court provided more specific instructions for the consolidated class action complaint, which was to contain “*no new products or causes of action; only those that have been already asserted in these cases*. None of the new ones on the proposed amended complaints; just the causes of action that are already existing and the products that are already existing.” 5/22/25 Hr’g. Tr. (ECF 80-1) at 20:19-23 (emphasis added). At the time, the NY Class Plaintiffs had sought to introduce the same expansive list of models included in the CCAC through a motion for leave to amend. *See* AG, ECF 405-02 (proposed amended complaint). The Court expressly denied including these new allegations in the consolidated complaint. *See* 5/22/25 Hr’g. Tr. (ECF 80-1) at 20:20-23 (instructing Class Plaintiffs that the consolidated class action complaint should include “[n]one of the new [products or causes of actions in] the proposed amended complaints.”) (emphasis added); *see also* ECF 60 at 2. The Court also explained why no additional products or claims would be allowed: “If I’m going to allow amended claims, new claims, new products, *it would be a never-ending process*.” 5/22/25 Hr’g. Tr. (ECF 80-1) at 25:24-25 (emphasis added).

Having instructed the parties to maintain, *and not expand*, the scope of the litigation, the Court recently issued a Scheduling Order setting forth the remaining key deadlines in the case. ECF 238.

C. The Consolidated Class Complaint Expands The Scope Of Class Plaintiffs' Claims Against Microsoft

Class Plaintiffs then filed a Consolidated Class Action Complaint against Microsoft and OpenAI on behalf of some of the existing named Plaintiffs in both New York and California. *See* CCAC. The CCAC adds new allegations that expand the case's scope.

CCAC's Expanded Class Definition (¶ 294)	CCAC's Expanded Direct Infringement Allegation (¶ 311)
<p>The Class consists of:</p> <p>All legal or beneficial owners of copyrighted works that: (A) were registered with the United States Copyright Office within five years of the work's first publication; (B) were downloaded or otherwise reproduced by OpenAI or Microsoft; (C) were registered with the United States Copyright Office before being downloaded or otherwise reproduced by OpenAI or Microsoft, or were registered within three months of first publication; and (D) are assigned one or more International Standard Books Number(s) (ISBN) or Amazon Standard Identification Number(s) (ASIN).</p>	<p>Defendants infringed on the rights, under 17 U.S.C. § 106, of Plaintiffs and members of the proposed Class by, among other things, reproducing the works owned by Plaintiffs and the proposed Class from shadow libraries and to create and develop datasets used to train their artificial intelligence models. Specifically, Plaintiffs and members of the Class assert claims for direct infringement against both Microsoft and OpenAI.</p>

As shown here, the proposed class definition set forth in paragraph 294 has no limitation on the purpose of the reproduction. CCAC, ¶ 294. Additionally, the charging allegation for the direct infringement claim in paragraph 311 contains no limitation on the scope of the reproduction, artfully using the qualifier “among other things” as well as the conjunction “and” before the critical clause about training, to sweep in a scope of alleged infringement that literally includes any reproduction of Plaintiffs' works. *Id.* at ¶ 311.

Taken together, these paragraphs have the effect of significantly expanding the scope of the class action against Microsoft. This combined language adds new claims arising from the alleged reproduction of Plaintiffs' works for any purpose regardless of the connection to the training of OpenAI's models. This would include Microsoft's training of LLMs independent of OpenAI, bringing into the case both historical activities of Microsoft Research and new activities of Microsoft undertaken long after the filing of the original NY Class Complaint.³ The language could also be read broadly enough to newly challenge Microsoft's decades-long operation of its search engine, Bing.

In addition to the changes wrought by paragraphs 294 and 311, the CCAC expands the claims as to Microsoft by adding in both older and newer released OpenAI GPT models and the products incorporating them that were not previously at issue in the NY Class Action as to Microsoft. In particular, paragraph 93 now alleges that GPT-4 *and the successor models* "[are] integrated into [Microsoft's] sales and marketing software, coding tools, productivity software, and cloud storage services." *Id.* at ¶ 93. This language (which was also the language included in the NY Class Plaintiffs' motion for leave to amend) would add the entire Microsoft suite of productivity applications to a case that has previously concerned the consumer Copilot chatbot f/k/a Bing Chat.

Finally, the new putative class definition stretches beyond books as the underlying copyrighted works at issue. By specifying that works with an ISBN *or* ASIN are included in the

³ During a meet and confer regarding the CCAC, Interim Lead Counsel for Plaintiffs indicated that Class Plaintiffs did not intend to include Microsoft-trained models within the scope of the litigation, but also refused to make any modification to the CCAC to reflect that supposedly more limited intended scope of claims. Declaration of Annette Hurst in Support of Microsoft's Motion to Strike Portions of the Consolidated Class Action Complaint ("Hurst Decl.") ¶ 6, Ex. A (6/20/25 J. Nelson email).

class definition, paragraph 294 would sweep into the scope of underlying works any item sold on Amazon. *Compare* CCAC ¶ 294 (including works that “are assigned one or more International Standard Books Number(s) (ISBN) *or Amazon Standard Identification Number(s) (ASIN)*”) (emphasis added); *with* NY Class Compl. ¶ 397 (defining “Nonfiction Class Works” as including texts that “have been assigned an International Standard Book Number (ISBN); and that fall within a Book Industry Standards and Communications (BISAC) code other than Reference (REF)”). This could include a wide variety of copyrighted content that has not previously been the subject of these cases, such as magazines, films, and music.

D. Class Plaintiffs File Two New Complaints Against Microsoft

Less than two weeks after filing the CCAC, certain former NY Class Plaintiffs filed a new putative class action complaint against Microsoft in the Southern District of New York alleging a single claim of direct infringement.⁴ *Bird, et al. v. Microsoft Corp.*, No. 1:2025-cv-05282 (S.D.N.Y.) (“*Bird*”), ECF 7 (“*Bird* Complaint”). The *Bird* Plaintiffs filed a Related Case Statement indicating that the new complaint is related to the MDL. *Bird*, ECF 3. While noting that the case differs from the MDL because it is “brought against Microsoft Corporation for direct copyright infringement,” as set out above, that is not really a distinction. Indeed, the proposed class definition in the *Bird* Complaint is identical to the one in this case except for its omission of OpenAI.

All legal or beneficial owners of copyrighted works that: (A) were registered with the United States Copyright Office within five years of the work’s first publication; (B) were downloaded or otherwise reproduced by Microsoft; (C) were registered with the United States Copyright Office before being downloaded or otherwise reproduced Microsoft, or were registered within three months of first publication; and (D) are assigned one

⁴ The CCAC dropped 27 plaintiffs named in the NY Class Action and CA Class Action. CCAC ¶¶ 21-42. A subset of the 19 dropped named plaintiffs from the NY Class Action are named in the *Bird* Complaint.

or more International Standard Books Number(s) (ISBN) or Amazon Standard Identification Number(s) (ASIN).

Id. ¶ 109. The existence of the *Bird* Complaint brought by many of the same named plaintiffs and the same counsel who previously were part of the NY Class Action seems clearly designed to hedge their bets in an implicit acknowledgment that the scope of the CCAC is broader than what was permitted by this Court.

Similarly, a week after the *Bird* Complaint, three new plaintiffs filed a putative class action complaint against Microsoft in the Northern District of California. *Denial, et al. v OpenAI, Inc., et al.*, No. 3:25-cv-05495 (N.D. Cal.) (“*Denial*”), ECF 1 (“*Denial* Complaint”). Although these plaintiffs were not named in the CA Class Action, they are represented by the same counsel as the CA Class Action Plaintiffs. *See id.* Counsel filed a Notice of Potential Tag-Along Action in the MDL for the *Denial* Complaint. Case MDL No. 3143, ECF 92.

The *Denial* Complaint alleges several claims against Microsoft and OpenAI, including direct and vicarious copyright infringement.⁵ *See Denial*, ECF 1, Count 1 and 2. The *Denial* Complaint includes two proposed classes, one for a “Non-Book Infringement Class” and another for “Unregistered Copyright Holders Class”:

All persons or entities domiciled in the United States that own a United States copyright in any textual work, where the work is registered with the United States Copyright Office, but are not assigned one or more International Standard Books Number(s) (ISBN) or Amazon Standard Identification Number(s) (ASIN).

Id. ¶ 84 (definition of “Non-Book Infringement Class”).

All persons or entities domiciled in the United States that own a United States copyright in any text data, who have not registered their works with the United States Copyright Office, including but not limited to books,

⁵ The *Denial* Plaintiffs also assert claims for unfair competition, violation of the California Comprehensive Computer Data Access and Fraud Act, violation of the Digital Millennium Copyright Act, conversion, unjust enrichment, breach of contract, violation of the Computer Fraud and Abuse Act, larceny, and violation of the Sherman Act. *See Denial*, ECF 1.

articles, essays, and other written works, that was accessed, copied, or used by OpenAI during the Class Period.

Id. ¶ 85 (definition of “Unregistered Copyright Holders Class”).

The *Bird* and *Denial* pleadings encompass the claims that were the subject of the pending motions for leave to amend and were explicitly discussed and disallowed at the May MDL organizational conference. There is no need for those claims to also be presented in this case.

ARGUMENT

Class Plaintiffs’ attempt to circumvent the Court’s Order and expand their claims against Microsoft through the CCAC should not be permitted, and their allegations expanding both the models and products at issue should accordingly be stricken pursuant to Federal Rules of Civil Procedure 12(f) and 16(f) and 28 U.S.C. § 1407.

District courts in this Circuit routinely reject “claims in amended complaints where the court granted leave to amend for a limited purpose and the plaintiff filed an amended complaint exceeding the scope of the permission granted.” *Palm Beach Strategic Income, LP v. Salzman*, 457 F. App’x 40, 43 (2d Cir. 2012); *see Sussman Sales Co., Inc. v. VWR Int’l, LLC*, No. 20 CIV. 2869 (JHR), 2025 WL 1004729, at *3 (S.D.N.Y. Apr. 1, 2025) (striking new claims added to amended complaint when plaintiff was granted leave to amend only to clarify the parties’ citizenship); *Starostenko v. UBS AG*, No. 19 CIV. 9993 (KPF), 2022 WL 1082533, at *3 (S.D.N.Y. Apr. 7, 2022) (striking new claims when “the Court ha[d] provided Plaintiffs with explicit and detailed instructions on what it would and would not permit them to include in their amended pleading”); *see also Cummings v. City of New York*, No. 19-CV-7723 (CM)(OTW), 2021 WL 1163654, at *3 (S.D.N.Y. Mar. 26, 2021) (disregarding allegations that exceeded the scope of the “specific purpose” the court had granted plaintiff leave to amend).

This MDL was consolidated and transferred to this court to “promote the just and

efficient conduct” of the underlying cases. 28 U.S.C. § 1407(a). Accordingly, this Court can “exercise the powers of a district judge in any district for the purpose of conducting pretrial. . . proceedings.” *Id.* § 1407(b); *see, e.g., In re Agent Orange Prod. Liab. Litig.*, 597 F. Supp. 740, 751-52 (E.D.N.Y. 1984), *aff’d sub nom. In re Agent Orange Prod. Liab. Litig. MDL No. 381*, 818 F.2d 145 (2d Cir. 1987) (finding that these powers include assuming “complete jurisdiction for pretrial purposes” and the “authority to decide all pretrial motions including dispositive motions.”).

In furtherance of the purpose of § 1407, and in accordance with Federal Rule of Civil Procedure 16, this Court issued a Case Management Order requiring, *inter alia*, Class Plaintiffs to file a consolidated class action complaint and explicitly limiting the scope of that complaint to the products and causes of action already at issue. *See* ECF 60 at 2. Failure to “obey a scheduling or other pretrial order,” including a case management order, comes with consequences. Fed. R. Civ. P. 16(f)(1)(C); *see also, e.g., Mahoney v. Yamaha Motor Corp. U.S.A.*, 290 F.R.D. 363, 371-72 (E.D.N.Y. 2013) (granting sanctions for failure to comply with case management order).⁶ This Court has a broad “inherent power and responsibility to manage its docket so as to achieve the orderly and expeditious disposition of cases.” *Huebner v. Midland Credit Mgmt., Inc.*, 897 F.3d 42, 53 (2d Cir. 2018) (cleaned up). This includes ridding the pleadings of allegations that defy the Court’s orders. *In re Worldcom, Inc. Sec. Litig.*, No. 03 CIV. 6225, 2004 WL 2889974, at *9-12 (S.D.N.Y. Dec. 15, 2004) (dismissing claims that ran afoul of MDL consolidation order); *Starostenko*, 2022 WL 1082533, at *3 (striking new claims that exceeded the scope of the court’s instructions for what could be included in the amended

⁶ *See also In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 460 F.3d 1217, 1252 (9th Cir. 2006) (“Failure to comply with case management orders in MDL proceedings... adversely affects the public interest, as well as the parties’ private interest, in expeditious resolution of litigation. A district court cannot manage its docket if such orders are not respected.”)

complaint).

Here, the CCAC violates the Court’s Case Management Order and other orders the Court has issued clarifying what is within the scope of this case. Class Plaintiffs include allegations that would vastly expand the scope of discovery that has been conducted over the past year and a half and which has been appropriately monitored and controlled by the Court so as to avoid undue burden, prejudice, or delay. This Court has expressed a desire for an efficient and streamlined resolution of these cases, and the CCAC impedes the Court’s ability to make that happen. Class Plaintiffs’ efforts to circumvent this Court’s specific instructions regarding the scope of the CCAC should not be allowed. Because the scope of this case is well-established, discovery is far along, and the Court provided specific instructions for what could (and could not) be included in the CCAC, these new expansive allegations should be stricken.

I. PORTIONS OF THE CCAC SHOULD BE STRICKEN BECAUSE THEY EXCEED THE SCOPE OF THE CASE SET BY THE COURT

The CCAC introduces three material expansions from the current case scope. *First*, the complaint no longer limits the allegations to infringement based on the training of OpenAI’s GPT models. *See* CCAC ¶¶ 294, 311. *Second*, Class Plaintiffs accuse older GPT models released before GPT-3 and more recent GPT models released after GPT-4 Turbo. *See id.* ¶¶ 4, 5, 81-82, 89-90, 93, 96-97, 117, 127, 161. *Third*, the new class definition no longer limits the affected works to books and sweeps in anything with an ASIN. *See id.* ¶ 294.

Each of these allegations would expand the existing scope of these cases in direct defiance of the Court’s Case Management Order. *See* ECF 60 at 2 (specifying that the consolidated Class Action Complaint was to “include only the same products and causes of action that have already been asserted in the pending putative class actions”). The Court may

accordingly strike these allegations. *Palm Beach Strategic Income*, 457 F. App'x at 43; *Sussman Sales*, 2025 WL 1004729, at *3; *In re Worldcom*, 2004 WL 2889974 at *9-12.

Specifically, Microsoft asks that the Court strike paragraphs 294 and 311 in their current form, and to strike references to GPT models and Microsoft products based on models released after GPT-4 Turbo and their “derivatives” and/or “successors,” as well as GPT-1 and GPT-2, in paragraphs 4, 5, 81-82, 89-90, 93, 96-97, 117, 127, and 161. The Court should order Class Plaintiffs to amend paragraphs 294 and 311 to limit the claims to reproduction of books in connection with the training of the permissible scope of OpenAI’s GPT models and Microsoft products.

A. Allegations Implicating Non-OpenAI-Trained Models Should Be Stricken Because Only OpenAI-Trained Models Have Been The Subject Of Discovery

Class Plaintiffs first seek to expand the case by, contrary to the prior complaints, not limiting the allegations to activities relating to the training of OpenAI’s GPT models. *See, e.g.*, CCAC ¶¶ 294, 311. The proposed class definition and related allegations that charge reproduction by Microsoft unrelated to OpenAI’s training of its models should be stricken because these cases—and over 18 months of discovery—have focused solely on OpenAI-trained models.⁷ Microsoft’s own LLM research work commenced in 2018—one year *before* any of the original conduct alleged as related to Microsoft’s relationship with OpenAI, which commenced with Microsoft’s first investment in OpenAI in 2019. Prior to that point, Microsoft’s LLM work was entirely independent of OpenAI. *See, e.g.*, AG, ECF 279 at 2.

Additionally, as explained at the technology tutorial, *see* 6/26/25 Hr’g. Tr. (ECF 303-1) at

⁷ Curiously, Class Plaintiffs only now seek to bring in Microsoft’s AI models. Indeed, most of Microsoft’s AI models were released well before the original allegations were raised, *e.g.*, Turing-NLG (released in 2020) and Megatron Turing (released in 2021). *See* AG, ECF 407 at 10, n.3. Seeking discovery into these LLMs at this stage of litigation is plainly unnecessary. There is no reason that Class Plaintiffs could not have taken issue with these LLMs at the outset.

270:9-272:23, in early 2024, long after this litigation started, Microsoft also acquired new technology and hired an entirely new internal team focused on Microsoft's development of its own commercial models. Microsoft continues to develop its own LLM work separate from OpenAI. *See, e.g., Introducing MAI-DS-R1*, Microsoft AI – Machine Learning Blog (Apr. 17, 2025), published at <https://techcommunity.microsoft.com/blog/machinelearningblog/introducing-mai-ds-r1/4405076>. Beyond Microsoft employees, there are also third parties that worked with Microsoft to co-develop AI technologies that would be implicated by such an expansion of the case.

None of this independent work has ever been the subject of this litigation. Indeed, Class Plaintiffs seem to recognize that these new theories are outside the scope of this case—hence the filing of the *Bird* Complaint where many of the named Plaintiffs who were formerly part of the NY Class Action case have decided to pursue separate claims against Microsoft. Lead Counsel for the Class Plaintiffs even offered on behalf of the *Bird* Plaintiffs that they would stay the *Bird* case pending resolution of the MDL class claims. Hurst Decl., ¶ 4. There is literally no reason for these different claims against Microsoft to proceed as part of a CCAC that has already been litigated vigorously for 18 months; they will only slow down resolution of the important fair use issue.

Microsoft does not understand why Class Plaintiffs refused to limit their pleading when they seemingly agree their purported claims against Microsoft related to its activity in training LLMs unrelated to OpenAI should be pursued in a separate action. In all events, the Court should strike paragraphs 294 and 311 and require Plaintiffs to amend them to include only claims against Microsoft related to its alleged involvement in OpenAI's training of its models.

B. Allegations Regarding GPT Models Released Before GPT-3 And After GPT-4 Turbo, And The Products Those Models Underlie, Should Be Stricken Because They Are Not Within The Scope Of Existing Discovery

The CCAC also directly contravenes the Court’s consolidation Order and related Case Management Order (and the Court’s prior orders prohibiting discovery into Microsoft’s new, post-Complaint products) by including GPT models that were not previously asserted against Microsoft. To date, Microsoft’s discovery obligations have been cabined to four of OpenAI’s LLMs: GPT-3, GPT-3.5, GPT-4, and GPT-4 Turbo.⁸ Class Plaintiffs now identify *four additional* LLMs—GPT-4V, GPT-4o, GPT-4o mini, and GPT-4.5—and also attempt to propound a catchall category including any of “the [LLMs’] successors.” According to the plain language of the CCAC, these additional “successors” also include at least the following *eleven* models: o1; o1-mini; o1-pro; o3; o3-mini; o3-pro; o4-mini; GPT-4.1; GPT-4.1 mini; GPT-4.1 nano; and GPT-image-1. In total, *fifteen new* LLMs are implicated as of today. Expanding the case to cover models *after* those previously at issue with respect to Microsoft could inherently sweep in Microsoft’s products that use those models—vastly expanding discovery.

As explained at the Tech Tutorial on June 26, 2025, expanding this case to encompass newer OpenAI LLMs (*i.e.*, post-GPT-4 Turbo) would add New Consumer Copilot to the case and as a result dramatically expand the scope of discovery:

⁸ OpenAI identified certain additional models in an interrogatory response even though they were not specifically identified in the formerly-operative complaint. *See* AG, ECF 281 at 1 (“This lawsuit, filed by book authors, centers around specific books datasets that were used to train GPT-3 and 3.5. OpenAI agreed to provide discovery into these and a number of other models—including GPT-1, GPT-2, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo—and, beyond that, to supplement its response to Interrogatory 11. . . .”) (cleaned up). Microsoft has never agreed these additional models were asserted as to it. Any documents produced by Microsoft related to these models were produced because they also contain material about models identified in the earlier, underlying complaints. Therefore, from Microsoft’s perspective, both the older (GPT-1 and GPT-2) and newer (post-GPT-4 Turbo) models remain outside the scope of the case as it relates to Microsoft.

New Consumer Copilot Is Different For Discovery Purposes

New Infrastructure

- New system architecture
- Built on different platform
- Output logs: entirely different system for capture and storage
- Incorporates only later-generation OpenAI LLMs (not previously in case as to Microsoft)

New Source Code/ Functions

- Entirely new code base stored in different location
- Plaintiffs have already spent ~50 days examining BingChat source code
- New capability to utilize numerous new external functions
- Many new app clients providing native support across multiple surfaces

New Custodians

- Built from scratch between May and October 2024
- Back-end code written by ~300 people
- Only 2 people overlap with BingChat

ECF 274 at 89.

Magistrate Judge Wang previously refused discovery into New Consumer Copilot, and ordered Microsoft to further explain the differences in the technology and to address the burden associated with adding that product to the case at the Technology Tutorial. Accordingly, as Microsoft explained, New Copilot is built on an entirely new system architecture, by an almost completely different group of people comprising nearly 300 custodians just on the technical side. *See* 6/26/25 Hr’g. Tr. (ECF 303-1) at 270:9-272:23. As another example, Plaintiffs had (at the time of the tutorial) already spent 50-person days looking at Bing Chat/M365 Copilot source code. *See* ECF 274 at 89. That effort would have to be started over entirely with respect to New Consumer Copilot.

Furthermore, the number of new products implicated by the allegations encompassing more recent GPT models is astronomical. Plaintiffs seek to include later OpenAI models integrated into Microsoft’s “sales and marketing software, coding tools, productivity software, and cloud storage services.” CCAC ¶ 93. ***This implicates hundreds of products and product***

features not previously in the case. The discovery necessary to address new products and features built on later OpenAI models would be orders of magnitude more than all discovery previously conducted.

These allegations are therefore directly contrary to the Court’s instructions that the CCAC should include *only the claims and products already asserted*. See 5/22/25 Hr’g. Tr. (ECF 80-1) at 39:24-40:2, 45:12-16 (responding to a question from the NY Class Plaintiffs to clarify that the models identified in the consolidated Class Action Complaint must “be in the case already” based on the “explicit models laid out” in the *Alter* case). Therefore, at the very least, the newer GPT models (*i.e.*, those developed after GPT-4 Turbo) are outside the scope of Class Plaintiffs’ claims properly at issue in the MDL and thus should be stricken. *In re Worldcom*, 2004 WL 2889974 at *9-12 (dismissing claims outside the scope of complaints underlying the MDL); *Sussman Sales*, 2025 WL 1004729 at *3 (striking new claims that were not within scope of court’s instructions for how to amend); *Starostenko*, 2022 WL 1082533 at *3 (similar).

Additionally, the earlier models (GPT-1 and GPT-2) were never part of the claims asserted against Microsoft (or OpenAI) in the NY Class Action, and thus cannot be fairly added against Microsoft now. It does not even make sense to include such theories against Microsoft. Those models were created before Microsoft’s investment in and business relationship with OpenAI. There is no factual explanation in the CCAC of how Microsoft was allegedly involved in those models; the only theory offered in the CCAC for Microsoft’s involvement in the training of OpenAI’s models is that it was an investor and business partner with OpenAI. See, e.g., CCAC ¶¶ 134-35, 316. Indeed, it is difficult to understand how Plaintiffs could allege, consistent with Rule 11, any involvement by Microsoft in connection with the training of those earlier models that predate Microsoft’s investment in OpenAI.

Microsoft respectfully asks the Court to strike the references to all models other than those released between GPT-3 and GPT-4 Turbo in paragraphs 4, 5, 81-82, 89-90, 93, 96-97, 117, 127, and 161 of the CCAC.

C. Allegations Regarding Non-Books Should Be Stricken Because The Litigation Has Focused Solely On The Use Of Books For Training OpenAI's LLMs

Finally, the class definition should be limited to ensure that it encompasses only the scope of works previously at issue in these cases—fiction and non-fiction books—and nothing more. These cases have never been about anything other than the purported infringement of books in the context of training OpenAI's models:

<u>NY Consolidated Class Complaint</u>	<u>CA Consolidated Class Complaint</u>
<p>The Fiction Author Class consists of: All natural persons in the United States who are the sole authors of, and legal or beneficial owners of Eligible Copyrights in, one or more Fiction Class Works; and all persons in the United States who are the legal or beneficial owners of Eligible Fiction Copyrights in one or more Fiction Class Works held by literary estates.</p> <p>AG, ECF 67 ¶ 393.</p> <p>Fiction Class Works is defined as follows: Any work of fiction, the text of which has been, or is being, used by Defendants to “train” one or more of Defendants’ large language models.</p> <p><i>Id.</i> ¶ 394.</p> <p>The Nonfiction Author Class consists of: All natural persons, literary trusts, and literary estates in the United States who are legal or beneficial owners of Eligible Nonfiction Copyrights in one or more Nonfiction Class Works; and all persons in the United States who are the legal or beneficial owners of</p>	<p>All persons or entities domiciled in the United States that own a United States copyright in any work that was used as training data for the OpenAI Language Models during the Class Period.</p> <p><i>In re OpenAI</i>, ECF 120 ¶ 53; <i>see also, id.</i> ¶ 22 (defining Infringed Works as the registrations included in Exhibit A to the Complaint, which consists solely of copyright registration numbers for books).</p>

<p>Eligible Nonfiction Copyrights in one or more Fiction Class Works held by literary estates.</p> <p><i>Id.</i> ¶ 396.</p> <p>Nonfiction Class Works is defined as follows: Any work of nonfiction (A) the text of which has been, or is being, used by Defendants to “train” one or more of Defendants’ large language models; (B) that have been assigned an International Standard Book Number (ISBN); and that fall within a Book Industry Standards and Communications (BISAC) code other than Reference (REF).</p> <p><i>Id.</i> ¶ 397.</p>	
--	--

The addition in CCAC paragraph 294 to any work with an ASIN number would expand the class beyond books to *any item sold on Amazon*. That allegation could expand this action to include content other than books such as magazines, films, and music. As drafted, the putative class definition is far broader than the prior class actions and thus in direct contravention of the Court’s Order not to expand the scope of the claims. Microsoft therefore respectfully requests that the reference to an ASIN number be stricken, or in the alternative if there is some reason why Class Plaintiffs require a reference to ASIN numbers in addition to the ISBN numbers that previously defined the scope of the works at issue, that the CCAC be amended to limit the relevant ASIN numbers to those encompassing fiction and non-fiction books only.

CONCLUSION

For the reasons presented herein, Microsoft respectfully requests the Court strike paragraphs 294 and 311 of the CCAC subject to limiting amendments as described herein, references in paragraphs 4, 5, 81-82, 89-90, 93, 96-97, 117, 127, and 161 to models other than those released between GPT-3 and GPT-4 Turbo, and reference to ASIN numbers in paragraph 294.

Dated: July 14, 2025

/s/ Annette L. Hurst

Annette L. Hurst (admitted *pro hac vice*)
ORRICK, HERRINGTON &
SUTCLIFFE LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105
Telephone: (415)773-5700
ahurst@orrick.com

Christopher Cariello
Marc Shapiro
ORRICK, HERRINGTON &
SUTCLIFFE LLP
51 West 52nd Street
New York: NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com
mrshapiro@orrick.com

Sheryl Koval Garko (admitted *pro hac vice*)
Laura Brooks Najemy (admitted *pro hac vice*)
ORRICK, HERRINGTON &
SUTCLIFFE LLP
222 Berkeley Street, Suite 2000
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 8801-1801
sgarko@orrick.com
lnajemy@orrick.com

Jeffrey S. Jacobson
FAEGRE DRINKER BIDDLE
& REATH LLP
1177 Avenue of the Americas
New York, New York 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted *pro hac vice*)
Kirstin L. Stoll-DeBell (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE

& REATH LLP
1144 Fifteenth Street, Suite 3400
Denver, CO 80202
Telephone: (303) 607-3500
jared.briant@faegredrinker.com
kirstin.stolldebell@faegredrinker.com

Carrie A. Beyer (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE & REATH LLP
320 South Canal Street, Suite 3300
Chicago, IL 60606-5707
Telephone: (312) 569-1000
Facsimile: (312) 569-3000
carrie.beyer@faegredrinker.com

Elizabeth M.C. Scheibel (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE & REATH LLP
2200 Wells Fargo Center, 90 S. 7th Street
Minneapolis, MN 55402
Telephone: (612) 766-7000
Facsimile: (612) 766-1600
elizabeth.scheibel@faegredrinker.com

Attorneys for Defendant Microsoft Corporation

PROOF OF SERVICE VIA ECF

On July 14, 2025, I caused the foregoing document to be served on all counsel of record via ECF.

/s/ Annette L. Hurst
Annette L. Hurst

CERTIFICATE OF WORD COUNT

Pursuant to Local Civil Rule 7.1, I hereby certify that the foregoing document was typed using 12-point, Times New Roman font and contains 5,751 words, exclusive of the case caption, table of contents, table of authorities, signature blocks, or any required certificates, but does include material contained in footnotes or endnotes.

/s/ Annette L. Hurst
Annette L. Hurst